REMARKS

On July 24, 2007, Applicants submitted a Request for Continued Examination. In the Office Action dated January 22, 2008, pending Claims 1-19 as submitted in the Supplemental Amendment dated October 5, 2007 were rejected. Claims 1, 10 and 19 are independent claims, while the remaining claims are dependent claims. Claims 1, 10 and 19 are being further amended herein to incorporate the subject matter of canceled Claims 4/5 and 13/14.

Rejections under 35 U.S.C. §103

Claims 1-3, 6-12, and 15-19 stand rejected under 35 USC § 103(a) as being unpatentable over U.S. Patent No. 6,904,402 to Wang et al. (hereinafter "Wang") in view of U.S. Patent No. 6,098,034 to Razin et al. (hereinafter "Razin") and further in view of *"Statistics-Based Segment Pattern Lexicon – A New Direction for Chinese Language Modeling"* by Yang et al. (hereinafter "Yang"). Claims 4-5 and 13-14 stand rejected under 35 USC § 103(a) as being unpatentable over Wang et in view of Razin and Yang and further in view of *"Color Set Size Problem with Applications to String Matching"* by Hui. Reconsideration and withdrawal of the present rejections are hereby respectfully requested, based upon the reasons set forth herein in addition to those set forth in the responses to previous office actions.

The present invention is directed to a method and apparatus for automatically extracting new words from a cleaned corpus, where the corpus can be in any language that may (or may not) have word boundaries (ranging from English or Latin to Chinese or Japanese). The instant invention segments a cleaned corpus to form a segmented corpus,

splits the segmented corpus to form sub strings, and counts the occurrences of each sub string appearing in the given corpus. Finally, the present invention filters out false candidates to output new words while reducing the memory requirements necessary to find the new words. Solely in an effort to expedite prosecution of the instant application, the independent claims have been amended to recite, *inter alia*, *"wherein the step of splitting and counting is implemented using a GAST contained in a reduced memory space ... wherein a GAST is implemented by limiting length of character sub strings"*. (*See* Claim(s) 1, 10 and 19; *see also* Specification at page 4, line 3 – page 6, line 11.)

The previously submitted comments regarding Wang remain equally applicable here. Additionally, Wang fails to teach, *inter alia*, "splitting" of the strings into substrings wherein *"the step of splitting and counting is implemented using a **GAST** contained in a reduced memory space ... "* as recited in the amended independent claims. Furthermore, the portion of Wang cited by the Examiner does not deal with "splitting" the corpus; it merely deals with "segmenting" the corpus using a Dynamic Order Markov Model (DOMM) data structure. (*See Wang* at col.8, lines 21-59.) As such, Wang does nothing to reduce the memory space required by a DOMM tree constructed from a large corpus other than providing extra memory. This stands in contrast to the instant invention, wherein the memory space problem is solved. (*See Specification* at page 4, line 3 - page 6, line 11.) Thus, Wang is not even aimed at solving a similar problem (i.e., finding new words and reducing the memory requirements necessary to find the new words). As a result, one skilled in the art would not be motivated to consult or modify Wang to achieve the instantly claimed invention.

Razin fails to overcome the deficiencies of Wang as set forth above. There is no suggestion or teaching in Razin that the segmenting and the splitting of the corpus is not dependent upon word boundaries. Razin also fails to teach, *inter alia*, "splitting" of the strings into sub-strings wherein *"the step of splitting and counting is implemented using a **GAST** contained in a reduced memory space ... wherein a GAST is implemented by limiting length of **character** sub strings"* as recited in the amended independent claims. In fact, Razin teaches away from this capability, teaching that the source text is segmented using a standard finite-state machine technique that recognizes patterns indicating **word and sentence** boundaries, and also that *"the atomic units represented at each node on the tree would be stemmed word packs, **not characters** or character strings"*. (*See Razin* at col. 11, lines 14-36 & col.12, line 55 – col.13, line 3.) Further, there is no suggestion or teaching that Razin discloses determining new words based upon the domain of the current corpus while reducing the memory requirements necessary to find the new words.

Yang fails to overcome the deficiencies of Wang and Razin as asserted above. Specifically, Yang also fails to teach, *inter alia*, "splitting" of the strings into sub-strings wherein *"the step of splitting and counting is implemented using a **GAST** contained in a reduced memory space ... "* as recited in the amended independent claims. Thus Yang does not teach or suggest the instantly claimed invention, and does not render it obvious either alone or in any combination with Wang or Razin.
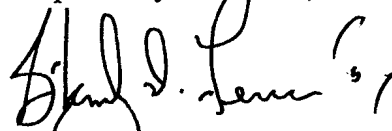
Hui cannot overcome the deficiencies of Wang, Razin and Yang as discussed above. Nothing in Hui renders the instantly claimed invention obvious, either alone or in any combination with Wang, Razin or Yang.

Applicants are not conceding in this application that the claims as amended and canceled herein are not patentable over the art cited by the Examiner, as the present claim amendments and cancellations are only for facilitating expeditious prosecution. Applicants respectfully reserve the right to pursue these and other claims in one or more continuations and/or divisional patent applications. Applicants specifically state that no amendment to any claim herein should be construed as a disclaimer of any interest in or right to an equivalent of any element or feature of the amended claim.

In view of the foregoing, it is respectfully submitted that independent Claims 1, 10 and 19 fully distinguish over the applied art and are thus allowable. By virtue of dependence from Claims 1 and 10, it is thus also submitted that Claims 2, 3, 6-9 and 11, 12, 15-18 are also allowable at this juncture. In summary, it is respectfully submitted that the instant application, including Claims 1-19, is presently in condition for allowance. Thus, Applicants respectfully request reconsideration and withdrawal of the outstanding claim rejections. Notice to that effect is hereby earnestly solicited.

Respectfully submitted,

Stanley D. Ference III
Registration No. 33,879

Customer No. 35195
FERENCE & ASSOCIATES LLC
409 Broad Street
Pittsburgh, Pennsylvania 15143
(412) 741-8400
(412) 741-9292 - Facsimile

Attorneys for Applicants